Final Report for Period: 08/2007 - 07/2008
Submitted on: 07/24/2008
Principal Investigator: Zwart, Albertus P.
Award ID: 0718701
Organization: GA Tech Res Corp - GIT
Submitted By:
Dai, Jiangang - Co-Principal Investigator
Title:
COLLABORATIVE RESEARCH: CSR---SMA: New Breakthrough in Analyzing Limited Resource Sharing Systems

Project Participants

Senior Personnel

Name: Zwart, Albertus
Worked for more than 160 Hours: Yes
Contribution to Project:

Collaborating with Dr. Jim Dai (Georgia Tech) and Ph.D. student Jiheng Zhang (Georgia Tech), Dr. Bert Zwart has studied limited processor sharing (LPS) queues when the jobsize distribution is not extreme. For an LPS queue with a general jobsize distribution, measure-valued fluid limits and diffusion limits are established. Furthermore, the steady-state of the diffusion process is justified to be a good approximation for the steady-state of the corresponding LPS queue. This line of research leads to three joint publications Zhang, Dai and Zwart (2007a), Zhang, Dai and Zwart (2007b), and Zhang and Zwart (2008).

Collaborating with Dr. Jim Dai (Georgia Tech), Dr. Mor Harchol-Balter (CMU), and Ph.D. student Varun Gupta (CMU), Dr. Bert Zwart has investigated LPS queues and many-server queues when jobsize distribution has a fixed, large squared coefficient of variation. It was demonstrated that the third moment of the jobsize can have a huge impact on the meaning waiting time of a job. One paper was submitted to a journal from this research.

Collaborating with Dr. G. Janssen and Dr. J. Van Leeuwaarden (Netherlands), Dr. Zwart has also focused on obtaining refinements of asymptotically optimal policies. For single class many server queues with exponential job sizes, new bounds and expansions have been developed for blocking and waiting probabilities. These new results have been applied to obtain the first mathematical proof that square root staffing yields staffing results that are at most one server away from the optimal number of servers. This has resulted in two publications.

Dr. Zwart was invited to present this research at the National Queueing Colloquium, Amsterdam, April 2008.

Name: Dai, Jiangang
Worked for more than 160 Hours: Yes
Contribution to Project:

Collaborating with Dr. Bert Zwart (Georgia Tech) and Ph.D. student Jiheng Zhang (Georgia Tech), Dr. Jim Dai has studied limited processor sharing (LPS) queues when the jobsize distribution is not extreme. For an LPS queue with a general jobsize distribution, measure-valued fluid limits and diffusion limits are established. These limits provide a solid foundation to develop approximations for various performance measures. This line of research leads to two joint publications Zhang, Dai and Zwart (2007a) and Zhang, Dai and Zwart

(2007b).

Collaborating with Dr. Bert Zwart (Georgia Tech), Dr. Mor Harchol-Balter (CMU), and Ph.D. student Varun Gupta (CMU), Dr. Jim Dai has investigated LPS queues and many-server queues when jobsize distribution has a fixed, large squared coefficient of variation. It was demonstrated that the third moment of the jobsize can have a huge impact on the meaning waiting time of a job. One paper was submitted to a journal from this research.

Collaborating with Dr. Tolga Tezcan (University of Illinois,Urbana-Champion), Dr. Jim Dai studied parallel server systems that consist of several job classes served by servers in server pools with different capabilities. A simple robust control policy is shown to minimize the total linear holding and reneging costs. This policy is asymptotically optimal under the many-server heavy traffic for parallel server systems when the service times are only server pool dependent and exponentially distributed. This research results in publications Tezcan and Dai (2007), and Dai and Tezcan (2008).

Dr. Dai gave a talk on this project in the Distinguished Lecture Series, Watson Research Center of IBM, April 2008.

Post-doc

Graduate Student

Name: Zhang, Jiheng

Worked for more than 160 Hours: Yes

Contribution to Project:

Jiheng Zhang is a Ph.D. student in the School of Industrial and Systems Engineering at Georgia Institute of Technology. This grant has been used to support Jiheng to conduct his thesis research on LPS queues.

Under the partial funding of this research grant, Jiheng Zhang has focused his research on limited processor sharing (LPS) queues when jobsize distribution is not extreme. Working with his advisors, Jim Dai and Bert Zwart at Georgia Tech, he has developed two-moment approximations for various performance measures such as average waiting time. These approximations use only the first two moments of the interarrival times and the jobsizes. These approximations are justified through a heavy traffic limit theorem, which guarantees that these approximations are asymptotically correct when the system becomes critically loaded and sharing level K becomes large.

Jiheng's research on LPS queues has generated three papers; the first was submitted to Mathematics of Operations Research, the second one was sumbitted to Annals of Applied Probability, and the third one was submitted to Queueing Systems. Jiheng presented his research at The INFORMS National Meeting, Seattle, November, 2007.

Jiheng is expected to receive his Ph.D. in 2008 academic year. The research projects from this grant have prepared him well for an academic position in US universities.

Undergraduate Student

Technician, Programmer

Other Participant

Research Experience for Undergraduates

Organizational Partners

Other Collaborators or Contacts

Tolga Tezcan, University of Illinois
Wuqin Lin, Northwestern University

Activities and Findings

Research and Education Activities:

Limited Processor Sharing (LPS) Queues

Motivated by applications in computer and communication systems, the PIs consider a processor sharing queue where the number of jobs being served is limited by K and the excess jobs wait in a buffer. Such a queue is called a limited processor sharing queue or an LPS queue.
Two distinct, but related, lines of research have been carried out for an LPS queue. The first line of research aims at an LPS queue whose jobsize distribution is not extreme, i.e., the third moment of the jobsize distribution is moderate, as compared with the variance.
The second line of research deals with the extreme jobsize distribution. When jobsize distribution is not extreme, the PIs have developed two-moment approximations for various performance measures such as average waiting time. These approximations use only the first two moments of the interarrival times and the jobsizes. These approximations are justified through a heavy traffic limit theorem, which guarantees that these approximations are asymptotically correct when the system becomes critically loaded and sharing level K becomes large. The heavy traffic limit theorem states that a certain stochastic process that keeps track of the system dynamics, when properly scaled, converges to a diffusion process, whose parameters are mapped from the first two moments of the jobsize and interarrival times distributions. Since the jobsize distribution is assumed to be general and a large number of jobs can simultaneously be served by the single server, we need to use a measure-valued process to keep track of the system dynamics. Very few prior work deal with diffusion approximations for measure-valued processes. In this research, the PIs carry out a complete program to prove the heavy traffic limit theorem for measure-valued processes. The PIs first study a measure-valued fluid model of an LPS queue and establish a fluid limit theorem. This work is documented in Zhang et al. (2007a). These fluid limits play an critical role in establishing a diffusion limit theorem, which is proved in Zhang et al. (2007b). To use the diffusion process for steady-state analysis of an LPS queue, one needs to justify the diffusion approximation continues to be valid in steady-state. Such a justification is carried out in Zhang and Zwart (2008); there numerical experiments are also carried out to confirm the accuracy of two-moment approximations.

Collaborating with CMU researchers, Dr. Mor Harchol-Balter and Ph.D. student Varun Gupta (CMU), the PIs have investigated LPS queues and many-server queues when jobsize distribution has a fixed, large squared coefficient of variation. It was demonstrated that the third moment of the jobsize can have a huge impact on the meaning waiting time of a job. The level of impact is also influenced by the availability of 'spare servers'. One paper was submitted to a journal from this research.

Findings: (See PDF version submitted by PI at the end of the report)

1) For a limited processor sharing (LPS) queue or a multi-server queue, the mean waiting can be sensitive to the third moment of the jobsize; thus, no two-moment approximation can be accurate for all jobsize distributions with a fixed first moment, and a fixed, large squared coefficient of variation (SCV).

2) For an LPS queue with moderate SCV, a two-moment approximation procedure is found accurate in predicting various performance measures such as average waiting time. The approximation is justified through

fluid limits and diffusion limits for measure-valued stochastic processes.

Training and Development:

Two Ph.D. students, Jiheng Zhang from Georgia Tech, and Varun Gupta from CMU, worked on the projects. The reserch involved here help them prepared for an academic position in US institutions.

Outreach Activities:

Journal Publications

Dai, JG; Hasenbein, JJ; Kim, B, "Stability of join-the-shortest-queue networks", QUEUEING SYSTEMS, p. 129, vol. 57, (2007). Published, 10.1007/s11134-007-9046-

Janssen, AJEM; Van Leeuwaarden, JSH; Zwart, B, "Gaussian expansions and bounds for the Poisson distribution applied to the Erlang B formula", ADVANCES IN APPLIED PROBABILITY, p. 122, vol. 40, (2008). Published,

Dai, J. G. and Lin, W., "Asymptotic optimality of maximum pressure policies in stochastic processing networks", Annals of Applied Probability, p. , vol. , (2008). Accepted,

Dai, J. G. and Tezcan, T., "Optimal control of parallel server systems with many servers in heary traffic", Queueing Systems, p. , vol. , (2008). Accepted,

Janssen, G., Van Leeuwaarden, J. and Zwart, B, "Refining square root staffing by expanding Erlang C", Operations Research, p. , vol. , (2008). Submitted,

Tezcan, T. and Dai, J. G., "Dynamic control of N-systems with many servers: asymptotic optimality of a static priority policy in heavy traffic", Operations Research, p. , vol. , (2008). Accepted,

Zhang, J., Dai, J. and Zwart, B., "Diffusion limits of limited processor sharing queues", Annals of Applied Probability, p. , vol. , (2007). Submitted,

Zhang, J., Dai, J. and Zwart, B., "Law of large number limits of limited processor sharing queues", Mathematics of Operations Research, p. , vol. , (2007). Submitted,

Zhang, J. and Zwart, B., "Steady state approximations of limited processor sharing queues in heavy traffic", Queueing Systems, p. , vol. , (2008). Submitted,

Books or Other One-time Publications

Web/Internet Site

Other Specific Products

Contributions

Contributions within Discipline:

The research results summerized in the finding section are all significant within discipline.

Contributions to Other Disciplines:

Contributions to Human Resource Development:

Contributions to Resources for Research and Education:

Contributions Beyond Science and Engineering:

Categories for which nothing is reported:

Organizational Partners
Activities and Findings: Any Outreach Activities
Any Book
Any Web/Internet Site
Any Product
Contributions: To Any Other Disciplines
Contributions: To Any Human Resource Development
Contributions: To Any Resources for Research and Education
Contributions: To Any Beyond Science and Engineering

For a limited processor sharing (LPS) queue or a multi-server queue, assume that the jobsize has a fixed first two moments. A new discovery was made that the mean waiting time can be sensitive to the third moment of the jobsize. When the squared coefficient of variation (SCV) of the jobsize distribution is high, the sensitivity is high. The practical implication is that no two-moment approximation will be accurate for all jobsize distributions with fixed first moment, and fixed, large SCV. The sensitivity also depends on the load of the system.

When the SCV is moderate and the system is heavily loaded, two-moment approximations have been developed for various performance measures such as average waiting time. Fluid limits and diffusion limits have been proved to justify these approximations. These limits deal with measure-valued stochastic processes. The methodologies have potential usage for other systems such as many-server parallel-server systems that model large-scale call centers.